

ORIGINAL ARTICLE

Assessing quality in inpatient psychiatry using outcome scales from the interRAI-Mental Health

Howard Barbaree*¹, Christopher Perlman², Elke Ham³, Gregory Brown⁴, Lynn Lightfoot⁵, John Hirdes²

¹Department of Psychiatry, University of Toronto, Toronto, ON, Canada

²School of Public Health Sciences, University of Waterloo, Waterloo, ON, Canada

³Waypoint Research Institute, Waypoint Centre for Mental Health Care, Penetanguishene, ON, Canada

⁴School of Criminal Justice, Nipissing University, North Bay, ON, Canada

⁵Lightfoot and Associates, Oakville, ON, Canada

Received: April 28, 2026

Accepted: June 11, 2026

Online Published: June 22, 2026

DOI: 10.63564/jha.v15n1p51

URL: <https://doi.org/10.63564/jha.v15n1p51>

ABSTRACT

Objective: This is the second study in a series to validate an outcome measure for inpatient psychiatry using outcome scales from the interRAI Mental Health (MH). To further validate the Composite Index of Inpatient Mental Health Status (CIIMHS) and its cross-cutting symptom domains by evaluating the relationship between the CIIMHS, its Domain Scores, and the Global Assessment of Functioning (GAF) (convergent validity). To evaluate the responsiveness of the interRAI MH scale scores and the CIIMHS and its Domain Scores to treatment effects. To use the CIIMHS to calculate an effect size (ES) for entry into a hospital's Balanced Scorecard (BSC).

Methods: We collected pre- and post-treatment scores on eight outcome scales from all RAI-MH assessments of hospital inpatients conducted over two fiscal years, along with GAF scores for one fiscal year, at a large mental health specialty hospital. Building on our previous work, which included a Principal Components Analysis (PCA) of the eight scales, yielding four standardized domain scores (psychosis, depression, impairment, and aggression), we combined these into the CIIMHS. To assess convergent validity, we calculated both parametric and nonparametric correlations between these measures and the GAF scores recorded during the same assessments. To evaluate construct validity, we analyzed changes in scale and domain scores from pre- to post-inpatient treatment across the hospital as a whole and separately for its acute care and addiction inpatient programs.

Results: Among all hospital patients and within the acute care program, pre- and post-treatment differences were statistically significant for seven of the eight RAI-MH scale scores, the composite measure, and the four domain scores (construct validity). Correlations between the GAF and the composite measure, as well as GAF and the impairment domain, approached or exceeded 0.60 (convergent validity). Correlations between GAF and other domains were smaller but were statistically significant, especially after treatment. The addiction program showed pre- and post-treatment differences in the scales with pre-treatment elevations, the composite measure, and the depression domain. Pre- and post-treatment outcome scales in the RAI-MH assessment platform demonstrated treatment effects, confirming good construct validity. Relationships with GAF indicated strong convergent validity. We suggest a methodology based on the CIIMHS to calculate a treatment ES for inclusion in the hospital's BSC. The CIIMHS is a common measure of illness severity.

Conclusions: T1vT2 differences, calculated quarterly, can be entered into hospital BSCs indicating quality of care.

Key Words: Inpatient psychiatry, Outcome measures, Health care quality, InterRAI Mental Health, Convergent validity, Construct validity, Acute care

*Correspondence: Howard Barbaree; Email: Howard.barbaree@utoronto.ca; Address: Department of Psychiatry, University of Toronto, ON, Canada.

1. INTRODUCTION

1.1 Quality of health care: The Donabedian Model

Modern discussions of hospital care quality have evolved from a conceptual framework proposed by Avedis Donabedian in a series of influential works published in the 1980s.^[1–3] According to Donabedian, the quality of care is proportional to the improvement in health status achieved by medical intervention, a definition that appears to have endured for the past 35 years.^[4] The model conceives of hospital care as comprising three components: Structure, Process, and Outcomes. Structure includes the hospital's infrastructure (physical plant and systems, equipment, budgetary resources, and hospital staff). Process includes all diagnostic, surgical, medical, and nursing procedures, as well as policies, procedures, and best practices. The third component of the model is patient outcome. Health status is a metric that quantifies a patient's health at a given point in time. If a patient's health status is assessed prior to the provision of care (T1) and after the provision of care (T2), the difference reflects the improvement in health due to treatment. According to the model, in high-quality clinical programs, improvements in structure and process lead to better outcomes. We refer to T1vT2 differences as treatment effects. We recognize that demonstrating treatment efficacy requires appropriate experimental controls, such as those used in Randomized Controlled Trials, which are not included in Quality Assessment.

1.2 Quality programs in inpatient psychiatry

A recent integrative review of peer-reviewed and industry-implemented Balanced Scorecards (BSCs)^[5] described their use in quality improvement programs across mental health settings worldwide. While hospitals have had no difficulty identifying critical structural and process elements for inclusion in their BSCs, they have struggled to identify meaningful outcomes. Outcome measurement in psychiatry has recently received more serious attention in the literature.^[6] An important distinction is made between patient-reported and clinician-reported outcome measures. Most BSC outcome entries have been patient-reported,^[5] emphasizing patients' and families' subjective experiences. The research host hospital's BSC contained multiple patient-reported measures. Additionally, clinician-reported outcomes have been limited in scope (e.g., % of patients with improved positive symptoms or changes in scores on traditional measures within a diagnostic category, such as depression). What has been lacking is an objectively quantified, clinician-scored overall measure of clinical severity and functioning.

1.3 Inpatient mental health status

Donabedian's model conceptualizes health status as best represented by a two-dimensional matrix, in which levels of function or performance are scaled along the vertical dimension and functional domains (physical, psychological) are categorized along the horizontal dimension. For example, cardiac health status might include domains such as blood pressure, lipid levels, smoking history, and other measurable factors. Overall health status would be represented by an average across these functional domains.

To fully implement the Donabedian model in inpatient psychiatry, we need to establish a conceptual framework for "inpatient mental health status" (IMHS) and devise a method to assess it. Which aspects of mental health are relevant, and how should they be evaluated? Psychiatric diagnoses are not particularly helpful for assessing outcomes, since most patients are discharged with the same diagnosis they had at admission. More important is the severity of mental health symptoms that led to admission. The DSM-5^[7] identified mental health symptoms relevant across psychiatric diagnoses. Known as cross-cutting symptoms, they are organized into 13 domains for adults, including depression, anger, mania, anxiety, somatic symptoms, suicidal thoughts, psychosis, sleep issues, memory, repetitive thoughts and behaviors, dissociation, personality functioning, and substance use. Only some of these (e.g., depression, anger, mania, suicidal thoughts) would be considered putative causes of hospital admission. IMHS could be calculated using standardized measures of these symptom domains, with IMHS defined as the mean score across these domains.

1.4 The RAI-MH

In 2005, the Ontario government mandated the use of the Resident Assessment Instrument-Mental Health^[8] across all publicly funded inpatient psychiatric services in Ontario, including acute, forensic, long-term, and geriatric psychiatry. Since then, the RAI-MH has been used in publicly funded inpatient psychiatric care in Ontario, administered at admission and discharge, and every three months for long-term patients, or whenever a significant change in clinical status occurs. By 2020, over 1.4 million assessments had been completed for more than 320,000 unique individuals across Canada.^[9] Previous researchers have successfully used the RAI-MH scales to evaluate outcomes.^[10,11]

An evaluation of the RAI-MH's performance was conducted at a large psychiatric hospital in Ontario, with findings reported in a 62-page internal report,^[12] and later in the *Canadian Journal of Psychiatry*.^[13] The authors provide extensive empirical data and observations on the instrument, along

with helpful guidance on its implementation and clinical application. However, they were skeptical of the RAI-MH outcome scales. They described the scale distributions as highly skewed, with very low baseline scores and minimal changes from admission to discharge. They doubted that the scales would be responsive to treatment effects, particularly for patients in acute care and addiction programs. Additionally, they reported low correlations between scale scores and the Global Assessment of Functioning (GAF),^[14, 15] suggesting poor convergent validity. Finally, they expressed disappointment, speculating that these scales would not provide a common measure of clinical severity and functioning across a heterogeneous clinical sample.

In a previous study,^[16] we used eight outcome scales from the RAI-MH assessment in a Principal Components Analysis (PCA) to derive scores across four outcome domains: psychosis, depression, impairment (including memory), and aggression (anger). These domain scores were combined to form the Composite Index of Inpatient Mental Health Status (CIIMHS). This measure was applied to the entire population ($N = 719$) of a large specialized mental health hospital in Ontario at two points: the beginning (T1) and end (T2) of a three-month hospital stay, or at admission and discharge for shorter stays. We evaluated content validity by comparing the scale items with a list of symptoms, complications, or functions suspected of contributing to hospital admission. To assess concurrent validity, we compared mean scores across hospital units with varying lengths of stay and levels of clinical complexity (acute, chronic, and complex chronic). Hospital discharge data and receiver operating characteristic (ROC) analysis were used to assess the CIIMHS's predictive validity. For construct validity, we examined changes in each measure from before to after treatment among short-term patients. Our findings provided strong evidence of validity across all four domains, leading us to conclude that the CIIMHS is a reliable and valid measure of inpatient mental health status.

The present study continues our validation of the CIIMHS, with an additional focus on its domain scores. First, we examined the relationship between the GAF and the CIIMHS and its domain scores as part of a test of convergent validity. Second, we examined changes from pre-treatment (T1) to post-treatment (T2) in each of the eight RAI-MH scale scores used in our previous paper. Third, we examined T1 vs. T2 (T1vT2) comparisons in the CIIMHS and its domain scores (psychosis, depression, impairment, and aggression), particularly in acute care and inpatient addiction programs (construct validity). Finally, we propose including the CIIMHS T1vT2 comparison in a hospital's BSC.

2. METHODS

2.1 Participants

Sample 1 included all patients ($N = 719$) receiving inpatient care at the participating hospital during one fiscal year. Scale scores were calculated at admission and discharge, or at the start and end of each 3-month quarter for long-stay patients. Subsample 1a ($n = 245$) participated in the hospital's acute care inpatient program. Subsample 1b ($n = 58$) participated in the hospital's inpatient addiction program. Data from Sample 1 were used to estimate treatment effects (scale score responsiveness and construct validity). Sample 2 included all patients ($N = 934$) who received inpatient hospital care during the second fiscal year. These data were used to calculate correlations between GAF and the CIIMHS and domain scores (convergent validity), separately by fiscal quarter, in a 4x-replication, non-independent groups design. Subsamples 2a-d ($n = 393, 385, 357, \text{ and } 344$) were assessed during the second fiscal year in quarters 1-4, respectively.

The patients' average age was 43 years. About 70% were male (73%, 68%). Most were never married (69%, 66%), with few married (14%, 17%) or having a live-in partner at the time of hospital admission. Additionally, 17% of both groups were widowed, separated, or divorced. Nearly all spoke English (99%). Although one-third of both groups graduated from high school and 23% had some post-secondary education, only a small percentage (8% and 14%) were employed, and 13% reported no income. The rest received a pension, social assistance, or disability benefits.

2.2 Setting

Waypoint Centre for Mental Health Care is a large psychiatric hospital (301 beds at the time) and a forensic mental health research facility in Penetanguishene, Ontario, Canada. It is one of four specialized mental health care centers in Ontario. The facility provides extensive acute and long-term inpatient and outpatient psychiatric services to the surrounding community, including Ontario's only high-security forensic mental health program.

2.3 Materials

The current study used eight outcome scales from the RAI-MH: the Aggressive Behavior Scale (ABS), the Activities of Daily Living (ADL) Hierarchy Scale, the Cognitive Performance Scale (CPS), the Depression Severity Index (DSI), the Instrumental Activities of Daily Living (IADL), the Positive Symptoms Scale (PSS), the Social Withdrawal Scale (SWS), and the Violence Sum (VS). As in our previous study,^[16] we selected these eight scales from the 15 available in the RAI-MH based on two criteria. First, the scale indicated a possible cause of inpatient hospital admission (symptoms,

complications, and functions). Second, the scale did not include static (historical) items, which could have reduced the scale's ability to reflect change from T1 to T2. A Principal Components Analysis identified four factors (domains): Depression, which combined the DSI and SWS; Impairment, which combined the CPS, ADL, and IADL; and Aggression, which combined the ABS and VS scales. The PSS, which was unrelated to any other scale at both T1 and T2, was standardized and represented the Psychosis domain. The three multi-scale domains were created by averaging the standardized scores of each scale within the respective factor. A composite measure was then created by summing the scores for the four domains. Please see our previous study for more details about these calculations.

The GAF is widely regarded as the most commonly used measure of impairment among patients with psychiatric disorders. It is a scale from 1 (lowest level of functioning) to 100 (highest level of functioning). The GAF was introduced in the DSM-III 17 as the fifth dimension of its multi-axial diagnostic system but was not included in the DSM-V.^[7]

2.4 Procedure

After completing the ethics review (Waypoint Centre for Mental Health Care Certificate #CRRA#12.03.01), Waypoint Decision Support provided anonymized data files containing item scores for the eight scales across all assessments for the two fiscal years. In Sample 2, the data were divided into four 3-month fiscal quarters. Separate analyses were conducted for each year (Sample 1; Sample 2).

The Sample 2 data file included GAF scores recorded at each assessment. At the end of each quarter, we identified all patients registered at the hospital and all patients discharged during that quarter. Seven hundred nineteen (Sample 1) and 934 (Sample 2) unique patients received inpatient care at Waypoint during the specified fiscal years. We then selected scale, domain, and CIIMHS scores for each patient at pre-treatment (T1) and post-treatment (T2) for each fiscal quarter, following the selection process described in detail in our previous study.

According to hospital protocol and practice, the outcome scales and the GAF were scored by different staff groups. RAI-MH outcome scales were scored by allied mental health professionals, mainly nurses, while the GAF was scored by medical staff, predominantly psychiatrists. Training and experience with the assessment instruments ensured reliable scoring. For the RAI-MH, studies of its psychometric properties have shown good reliability and validity.^[18,19] Multidisciplinary allied health professionals were trained in its use.

The Hospital's Decision Support conducted studies of interrater agreement in the scoring of the RAI-MH outcome scales and found agreement among hospital raters to be "good-to-excellent." For the GAF, instructions for use and scoring, including described anchor points, are presented in the DSM-III, and practicing psychiatrists have had both training and extensive experience in its use.

2.5 Statistical analysis

To assess convergent validity, we examined correlations between the GAF and measures derived from RAI-MH scale scores. Given skewness in these distributions, we calculated both Pearson's *R* and Spearman's *Rho*.

To evaluate scale-score responsiveness, we performed paired *t*-tests comparing T1 and T2 for the eight RAI-MH scale scores in Sample 1. Additionally, using the CIIMHS and domain scores, we conducted one-way within-subjects ANOVAs comparing T1 and T2 across the entire sample (Sample 1) and separately for the Acute Care program (Sample 1a) and the Addictions program (Sample 1b).

The single-factor within-subjects ANOVA provides three estimates of variance (Sums of Squares): (1) the difference between the means at T1 and T2 (Effect); (2) differences among patients' marginal scores; and (3) error (see Equation 1).

$$SS_{Total} = SS_{Effect} + SS_{Patients} + SS_{Error} \quad (1)$$

The variance estimate for error reflects the consistency of change from T1 to T2 across patients. The more similar the change is from T1 to T2 across patients, the smaller the error. We compared programs' T1vT2 effect sizes (ESs). The ES is calculated as Equation 2:

$$Partial\ Eta\ Squared(\eta_p^2) = SS_{Effect} / (SS_{Effect} + SS_{Error}) \quad (2)$$

Therefore, two factors determine the ES. The larger the difference between the means at T1 and T2 and the more consistent the T1-to-T2 changes across patients, the larger the ES. This makes Partial Eta Squared (η_p^2) an ideal statistical model for assessing quality of care as defined by the Donabedian model.

3. RESULTS

Table 1 presents results related to convergent validity. Overall, the CIIMHS and all four domains showed significant correlations with the GAF. For Pearson *R*, the CIIMHS and the Impairment Domain were strongly correlated with the

GAF (the mean *R* at T1 for the Composite was -0.414; at T2 was -0.546; the mean *R* for the Impairment Domain at T1 was -0.533; and at T2 was -0.665). Spearman's Rho results confirmed the Pearson *R* findings. Most other domain scores (Aggression, Psychosis, and Depression) also showed sig-

nificant correlations with the GAF, especially at T2, though with lower *R* values. The mean *R* for Aggression was -0.223 at T1 and -0.327 at T2. The symptom domain's correlations (Psychosis and Depression) were statistically significant but had lower *R* values.

Table 1. Correlations between the GAF and the CIIMHS and its Components using four overlapping non-independent sub-samples (Sample 2; *N* = 934)

Samples by Quarters	Measures	Correlations			
		Pearson's <i>R</i>		Spearman's Rho	
		T1	T2	T1	T2
Subsample 2a (<i>n</i> = 393)	Composite	-0.379**	-0.518**	-0.398**	-0.538**
	Impairment	-0.537**	-0.633**	-0.421**	-0.620**
	Aggression	-0.192**	-0.305**	-0.187**	-0.323**
	Psychosis	-0.046	-0.150**	-0.078	-0.216**
	Depression	-0.021	-0.151**	-0.008	-0.222**
Subsample 2b (<i>n</i> = 385)	Composite	-0.409**	-0.536**	-0.418**	-0.540**
	Impairment	-0.550**	-0.645**	-0.482**	-0.594**
	Aggression	-0.245**	-0.284**	-0.238**	-0.334**
	Psychosis	-0.039	-0.184**	-0.077	-0.278**
	Depression	-0.129*	-0.263**	-0.146**	-0.302**
Subsample 2c (<i>n</i> = 357)	Composite	-0.394**	-0.551**	-0.378**	-0.566**
	Impairment	-0.568**	-0.683**	-0.493**	-0.651**
	Aggression	-0.185**	-0.314**	-0.213**	-0.363**
	Psychosis	-0.041	-0.223**	-0.091	-0.318**
	Depression	-0.082	-0.198**	-0.153**	-0.269**
Subsample 2d (<i>n</i> = 344)	Composite	-0.475**	-0.580**	-0.453**	-0.595**
	Impairment	-0.610**	-0.698**	-0.532**	-0.631**
	Aggression	-0.271**	-0.404**	-0.308**	-0.425**
	Psychosis	-0.135*	-0.220**	-0.184**	-0.299**
	Depression	-0.001	-0.255**	-0.077	-0.318**

Note. **p* < .05; ***p* < .01

Table 2 presents paired *t*-test results comparing pre- and post-treatment scores on the eight RAI-MH outcome scales. The top panel shows results for the entire hospital (Sample 1; *N* = 719), the middle panel presents results for the acute care program (Subsample 1a; *n* = 241), and the bottom panel presents results for the addictions program (Subsample 1b; *n* = 59). Across the hospital and among acute care patients, seven scales showed statistically significant reductions from pre- to post-treatment. Changes on the ADL scale were not significant, likely due to the very low pre-treatment scores. Among addiction patients, only the DSI and SWS scales had elevated pre-treatment scores and showed statistically significant decreases from pre- to post-treatment.

Table 3 presents the results of a one-way within-subjects

ANOVA comparing T1 and T2 on the composite score (CIIMHS) and the four outcome domains (domain scores). The top panel displays results for the entire hospital. Score reductions from T1 to T2 were statistically significant for both the CIIMHS and all four domains. Across the hospital, as expected, the CIIMHS decreased from T1 (mean = 5.63) to T2 (mean = 4.34), $F(1, 718) = 95.02, p < .001; \eta_p^2 = 0.117$, indicating a moderate ES. The hospital also showed statistically significant decreases from T1 to T2 in all four mental health domains: Psychosis, $F(1, 241) = 21.17, p < .001$; Depression, $F(1, 241) = 40.42, p < .001$; Impairment, $F(1, 241) = 53.33, p < .001$; and Aggression, $F(1, 241) = 27.61, p < .001$.

The middle panel presents results for the acute care program.

The program’s CIIMHS score decreased from T1 (3.44) to T2 (1.54), $F(1, 241) = 75.54, p < .001, \eta_p^2 = 0.235$, indicating a large ES. Additionally, the program showed statistically significant reductions from T1 to T2 across all four mental health domains: Psychosis, $F(1, 241) = 28.88, p < .001$; Depression, $F(1, 241) = 53.57, p < .001$; Impairment, $F(1, 241) = 24.97, p < .001$; and Aggression, $F(1, 241) = 11.80, p < .001$.

The bottom panel displays the results for the addiction program. A significant decrease in the mean CIIMHS score from T1 (2.78) to T2 (1.14) was observed, $F(1, 58) = 20.55, p < .001, \eta_p^2 = 0.262$, indicating a large ES. Significant decreases were also observed in the Depression domain from T1 to T2, $F(1, 58) = 24.51, p < .001$.

Table 4 outlines how results from these analyses might be entered into a hospital’s BSC. The table presents data for 4 consecutive fiscal quarters in a single year for four non-independent groups (these numbers include all patients in the hospital during each quarter). The groups are not independent because many patients are hospitalized for more than one quarter. Means for the CIIMHS at T1 and T2 are presented for each quarter, along with the T1vT2 difference. ESs are presented separately for each quarter. Rolling ESs are then calculated as the mean of the current ES and each previous ES. The final entry in that row is the Rolling Annual ES, calculated quarterly. We suggest that this statistic be entered into the hospital’s BSC each quarter. Table 4 also includes the same GAF calculations for comparison.

Table 2. Paired *t*-tests comparing T1vT2 means using 8-outcome scales from the RAI-MH

Scale	T1 Mean	T1 SD	T2 Mean	T2 SD	<i>t</i>	<i>df</i>	<i>p</i>
Whole Hospital (Sample1)							
ABS	1.31	2.38	0.95	2.04	4.17	718	< .001
CPS	0.82	1.44	0.71	1.36	3.05	718	.002
DSI	1.62	2.40	0.92	1.71	7.49	718	< .001
VS	3.14	3.74	2.55	3.41	4.52	718	< .001
PSS	2.52	3.31	1.64	2.77	7.51	718	< .001
IADL	6.50	8.34	6.03	8.83	3.14	718	.002
SWS	2.32	3.56	1.65	3.14	4.76	718	< .001
ADL _(Hierarchy)	0.39	0.98	0.37	1.02	0.86	718	ns
Acute Care (Subsample 1a)							
ABS	0.89	1.86	0.51	1.42	2.89	240	.004
CPS	0.25	0.81	0.02	0.16	4.40	240	< .001
DSI	1.77	2.36	0.61	1.39	6.44	240	< .001
VS	1.78	3.56	1.13	2.75	2.81	240	.005
PSS	1.59	2.55	0.56	1.58	6.21	240	< .001
IADL	1.76	2.62	1.09	2.70	3.67	240	< .001
SWS	2.24	3.17	0.74	1.95	6.59	240	< .001
ADL _(Hierarchy)	0.10	0.55	0.07	0.44	0.80	240	ns
Addictions (Subsample 1b)							
ABS	0.05	0.39	0.14	0.63	-1.09	58	ns
CPS	0.02	0.13	0.00	0.00	1.00	58	ns
DSI	3.69	3.69	1.20	2.35	5.14	58	< .001
VS	0.53	1.15	0.46	1.16	0.68	58	ns
PSS	0.88	1.68	0.41	1.30	1.90	58	ns
IADL	0.42	1.53	0.27	0.87	0.89	58	ns
SWS	2.95	3.75	0.69	1.55	4.21	58	< .001
ADL _(Hierarchy)	0.02	0.13	0.00	0.00	1.00	58	ns

Table 3. Comparing CIIMHS and domain scores at T1vT2 for the whole hospital and 2 acute programs

		Mean	SD	Mean	SD	F	df	sig	Partial η^2
		at T1	at T1	at T2	at T2				
Whole Hospital (Sample 1)									
N = 719	CIIMHS	5.63	4.63	4.34	4.63	95.02	1,718	< .001	0.117
	Psychosis	6.63	8.71	5.12	8.47	21.17	1,718	< .001	0.029
	Depression	5.93	7.14	4.05	5.97	40.42	1,718	< .001	0.053
	Impairment	7.64	11.36	6.32	10.50	53.33	1,718	< .001	0.069
	Aggression	7.98	9.68	6.23	8.78	27.61	1,718	< .001	0.037
Acute Care (Subsample 1a)									
n = 241	CIIMHS	3.44	3.12	1.54	2.55	73.54	1,240	< .001	0.235
	Psychosis	4.20	6.71	1.73	4.93	28.88	1,240	< .001	0.107
	Depression	6.06	6.86	2.20	4.74	53.57	1,240	< .001	0.182
	Impairment	2.14	4.41	0.84	2.44	24.97	1,240	< .001	0.094
	Aggression	4.81	8.47	2.95	6.59	11.80	1,240	< .001	0.047
Addictions (Subsample 1b)									
n = 59	CIIMHS	2.78	2.42	1.14	2.05	20.55	1,58	< .001	0.262
	Psychosis	2.32	4.43	1.26	4.07	2.25	1,58	ns	0.037
	Depression	10.26	10.04	3.28	6.30	24.51	1,58	< .001	0.297
	Impairment	0.34	1.30	0.98	2.37	1.65	1,58	ns	0.028
	Aggression	0.98	2.37	1.05	2.90	<1.00	1,58	ns	0.001

Table 4. T1vT2 differences between means, effect sizes, and rolling effect size for CIIMHS and GAF over 4 sequential fiscal quarters for acute care

Fiscal Quarter	1	2	3	4
Subsample	2a	2b	2c	2d
n	88	91	90	92
CIIMHS				
Mean T1	3.39	3.01	3.80	3.85
Mean T2	2.09	2.07	1.93	1.41
T1vT2 Difference	1.31	0.94	1.87	2.44
Effect Size	0.127	0.071	0.204	0.235
Rolling ES	0.127	0.099	0.134	0.159
GAF				
Mean T1	39.88	39.52	38.22	40.44
Mean T2	44.97	44.73	43.17	44.52
T1vT2 Difference	-5.09	-5.21	-4.95	-4.08
Effect Size	0.227	0.194	0.199	0.209
Rolling ES	0.227	0.211	0.207	0.207

Note. CIIMHS = Composite Index of Inpatient Mental Health Status; GAF = Global Assessment of Functioning

4. DISCUSSION

In our previous study,^[16] we used eight outcome scales from the RAI-MH assessment to derive scores across four outcome

domains: psychosis, depression, impairment, and aggression. These domain scores were combined to form the CIIMHS. We evaluated the CIIMHS's content, concurrent, predictive,

and construct validity and found strong evidence of validity across all four spheres. In the current study, we found high correlations between GAF and the CIIMHS, particularly its Impairment domain, indicating strong convergent validity. Additionally, we confirmed the RAI-MH scales' responsiveness to treatment in acute and addiction programs, addressing previous concerns. The CIIMHS combines 40+ RAI-MH items, scored by professional staff with the most comprehensive knowledge of the individual patient, gained through direct observation, chart review, and case conferences. These 40+ items are combined into 8 scales that cover a broad range of issues related to the causes of hospital inpatient admission. These scales are combined into 4 cross-cutting symptom domains, which in turn combine to form the validated CIIMHS. RAI-MH has been mandated in all psychiatric inpatient settings in Ontario since 2005, and the implementation of this instrument in quality improvement efforts should thereby be facilitated. Adding CIIMHS to a hospital's BSCs involves analyzing existing data for both current and historical evaluations.

Improvements to the CIIMHS are planned. In the present study, the ADL hierarchy scale was found to be unresponsive to treatment effects in this population. Removing that item yields a modified CIIMHS that correlates with the original at $R = 0.94$. Going forward, the CIIMHS will be based on the 7 scales that were responsive to treatment. As we reported in our original study, the CIIMHS is incomplete because we were unable to identify dynamic RAI MH items covering self-harm and substance abuse. Further work is planned to develop such items. We propose that the CIIMHS could be enhanced by developing and adding such scales. Although still incomplete, the CIIMHS, with its four cross-cutting symptom domains, provides a standard measure of illness severity among inpatient psychiatric patients and demonstrates strong content, concurrent, predictive, and construct validity.

The current study demonstrates how the CIIMHS Effect Size, derived from the within-subjects ANOVA of the CIIMHS T1vT2 comparison, can be incorporated into a Hospital's Balanced Scorecard. While the ES varies from quarter to quarter, using a rolling ES that includes data from the most recent 4 quarters yields a more stable outcome. The "Annual ES Calculated Quarterly" ESs quantify the magnitude of a statistical difference between groups, independent of sample size. It can be expressed in terms of the measure used (CIIMHS differences) or as a standardized measure.^[20] We have used the standardized measure of ES commonly derived from ANOVA, namely η_p^2 .^[21] This measure of ES can be understood as the proportion of the total variance in CIIMHS that is accounted for by the statistical effect (the

difference between groups). Cohen^[22] provided categories of ESs, which we will report here as our empirically derived validity coefficients. Categories for η_p^2 are as follows: 0.0099 to 0.0587 is small; 0.0588 to 0.1378 is medium; and above 0.1379 is large. A study reported in the *Journal of Counseling Psychology*, 23 calculated values of η_p^2 from 11,044 inferential statistics reflecting treatment ESs from 1970 to 1979 and found a median value of 0.0830.

These results support the use of the CIIMHS and its domain scores to assess inpatient mental health status and quality of care within the Donabedian quality framework. Other cross-cutting domains from the DSM-5 list might also warrant consideration for inclusion. However, these additions require proper validation. Meanwhile, the current CIIMHS provides a suitable measure for outcome evaluation and research, serving as a common metric for clinical severity and functioning.

Amid developments such as deinstitutionalization, significant investment in community programs, the rise of the recovery movement, and significant budget pressures, inpatient psychiatry has been relegated to a critical yet limited role within the continuum of psychiatric care.^[24] Extreme financial pressures have made safety and crisis stabilization its primary focus. Hospital crisis admissions are essential to prevent harm, and the main goals of inpatient care are to address these functional areas—such as severe symptoms and risk-related complications or impairments—to ensure a safe transition back to the community. In this context, we conceptualize inpatient mental health status as determined by cross-cutting symptom domains that necessitate hospitalization and later prevent discharge. The quality of inpatient mental health care should be directly related to how effectively treatment ameliorates these symptoms (complications, impairments) and facilitates successful and rapid discharge to the community.

ACKNOWLEDGEMENTS

Waypoint's Senior Leadership Team has recognized the importance of outcome measures and has provided unwavering support for the research described here. The project was fully funded by Waypoint resources. We express our sincere gratitude to Debra Wicks and her team in Decision Support, particularly Mike Radko, for providing quarterly RAI-MH data. Dr. Peter Prendergast provided helpful comments on an earlier draft. Finally, we thank our interRAI colleagues, especially the late Brant Fries, who have provided advice, constructive criticism, and support throughout our work on the project. Author HB used Grammarly for grammar and spell check.

AUTHORS CONTRIBUTIONS

HB: Conceptualization, data analysis, internal funding, literature review, methodology, project administration, writing the first draft, review & editing. CP: Ethics approval, conceptualization, methodology, review & editing. EH: Data curation, writing, review & editing, methodology. GB: Conceptualization, writing, review & editing. LL: Conceptualization, review & editing. JH: Conceptualization, methodology, resources, writing, review & editing.

FUNDING

The author(s) declare that the research was conducted with internal hospital support. No external financial support was received for the research or for the publication of this article.

CONFLICTS OF INTEREST DISCLOSURE

The authors declare they have no conflicts of interest.

INFORMED CONSENT

Written informed consent for participation was not required from participants or their legal guardians/next of kin. We received an anonymized data file after most patients had left the hospital, and we had no contact with participants as part of this research.

ETHICAL STATEMENT

The studies involving human participants described in this manuscript were reviewed by Waypoint's Research Ethics Board. Approval of the application titled "The Development of Clinical Outcome Measures based on the

interRAI-MH among clinical programs at Waypoint Center for Mental Health Care" was documented in Certificate #CRR#12.03.01. The studies were conducted in accordance with local legislation and institutional requirements.

ETHICS APPROVAL

The Publication Ethics Committee of the Association for Health Sciences and Education. The journal's policies adhere to the Core Practices established by the Committee on Publication Ethics (COPE).

PROVENANCE AND PEER REVIEW

Not commissioned; externally double-blind peer reviewed.

DATA AVAILABILITY STATEMENT

The datasets generated and/or analyzed during the current study would be made available from the corresponding author upon reasonable request.

DATA SHARING STATEMENT

The data can be requested from the corresponding author.

OPEN ACCESS

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0/>).

COPYRIGHTS

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

REFERENCES

- [1] Donabedian A. The quality of care: how can it be assessed? *JAMA*. 1988; 260(12): 1743-1748. PMID: 3045356. <https://doi.org/10.1001/jama.1988.03410120089033>
- [2] Donabedian A. An introduction to quality assurance in health care. New York: Oxford University Press; 2003. <https://doi.org/10.1093/oso/9780195158090.001.0001>
- [3] Donabedian A, Wheeler JRC, Wyszewianski L. Quality, cost, and health: an integrative model. *Med Care*. 1982; 20(10): 975-992. PMID: 6813605. <https://doi.org/10.1097/00005650-198210000-00001>
- [4] Agency for Health Care Research and Quality. Health care quality [Internet]. Accessed January 24, 2019. Available from: <https://www.ahrq.gov/topics/quality.html>
- [5] Brimelow RE, Amalathas A, Beattie E, et al. The use of balanced scorecards in mental health services: an integrative review and thematic analysis. *J Behav Health Serv Res*. 2023; 50(1): 128-146. PMID: 35835954. <https://doi.org/10.1007/s11414-022-09806-3>
- [6] Ryland H, Carlile J, Kingdon D. A guide to outcome measurement in psychiatry. *BJPsych Adv*. 2021; 27: 263-271. <https://doi.org/10.1192/bja.2020.58>
- [7] American Psychiatric Association. Diagnostic and statistical manual of mental disorders, 5th ed. Arlington: American Psychiatric Publishing; 2013. 733-748 p. <https://doi.org/10.1176/appi.books.9780890425596>
- [8] Hirdes J, Ljunggren G, Morris J, et al. Reliability of the interRAI suite of assessment instruments: a 12-country study of an integrated health information system. *BMC Health Serv Res*. 2008; 8: 277. PMID: 19115991. <https://doi.org/10.1186/1472-6963-8-277>
- [9] Hirdes JP, van Everdingen C, Ferris J, et al. The interRAI suite of mental health assessment instruments: an integrated system for the continuum of care. *Front Psychiatry*. 2020; 10: 926. PMID: 32076412. <https://doi.org/10.3389/fpsy.2019.00926>
- [10] Perlman CM, Hirdes JP, Barbaree HE, et al. Development of mental health quality indicators (MHQIs) for inpatient psychiatry based on the interRAI mental health assessment. *BMC Health Serv Res*. 2013; 13: 15. PMID: 23305286. <https://doi.org/10.1186/1472-6963-13-15>

- [11] Tolonen J, Jaaskelainen E, Leppanen V, et al. The impact of psychiatric rehabilitation: a study of outcomes of persons with severe mental disorders. *Psychiatr Fenn*. 2022; 53: 204-219.
- [12] Urbanoski KA, Rush B, Ehtesham S. The Resident Assessment Instrument at CAMH: a comprehensive review of current practices and potential opportunities. Toronto: Centre for Addiction and Mental Health; 2009.
- [13] Urbanoski K, Mulsant B, Willet P, et al. Using routinely collected clinical assessments in mental health services: the Resident Assessment Instrument-Mental Health. *Can J Psychiatry*. 2014; 59(7): 687-695. PMID: 25165782. <https://doi.org/10.1177/070674371405900709>
- [14] Moos RH, McCoy L, Moos BS. Global assessment of functioning (GAF) ratings: determinants and role as predictors of one-year treatment outcomes. *J Clin Psychol*. 2000; 56: 449-461. [https://doi.org/10.1002/\(SICI\)1097-4679\(200004\)56:4<449::AID-JCLP1>3.0.CO;2-8](https://doi.org/10.1002/(SICI)1097-4679(200004)56:4<449::AID-JCLP1>3.0.CO;2-8)
- [15] Moos RH, Nichol AC, Moos BS. Global assessment of functioning ratings and the allocation and outcomes of mental health services. *Psychiatr Serv*. 2002; 53: 730-737. PMID: 12045311. <https://doi.org/10.1176/appi.ps.53.6.730>
- [16] Barbaree HE, Perlman C, Ham E, et al. Validation of a composite outcome measure for inpatient psychiatry using scales from the interRAI-MH. *Front Psychiatry*. 2025; 16: 1486734. PMID: 39980981. <https://doi.org/10.3389/fpsy.2025.1486734>
- [17] American Psychiatric Association. Diagnostic and statistical manual of mental disorders, 3rd ed. Arlington: American Psychiatric Publishing; 1987.
- [18] Hirdes JP, Marhaba M, Smith TF, et al. Development of the Resident Assessment Instrument-Mental Health (RAI-MH). *Hosp Q*. 2000; 4(2): 44-53. PMID: 11484623. <https://doi.org/10.12927/hcq.2000.16756>
- [19] Hirdes JP, Smith TF, Rabinowitz T, et al. The Resident Assessment Instrument-Mental Health (RAI-MH): inter-rater reliability and convergent validity. *J Behav Health Serv Res*. 2002; 29: 419-432. PMID: 12404936. <https://doi.org/10.1007/BF02287348>
- [20] Cohen J. The earth is round ($<.05$). *Am Psychol*. 1994; 49(12): 997-1003. <https://doi.org/10.1037/0003-066X.49.12.997>
- [21] Richardson JTE. Eta squared and partial eta squared as measures of effect size in educational research. *Educ Res Rev*. 2011; 6: 135-147. <https://doi.org/10.1016/j.edurev.2010.12.001>
- [22] Cohen J. Statistical power analysis for the behavioral sciences. New York: Academic Press; 1969. 278-280 p.
- [23] Haase RF, Waechter DM, Solomon GS. How significant is a significant difference? Average effect size of research in counseling psychology. *J Couns Psychol*. 1982; 29: 58-65. <https://doi.org/10.1037/0022-0167.29.1.58>
- [24] Glick D. Inpatient psychiatric care in the 21st century: the need for reform. *Psychiatr Serv*. 2011; 62(2): 206-209. PMID: 21285100. https://doi.org/10.1176/ps.62.2.pss6202_0206